ASSIGNMENT No. 02

Statistics for Management (1430) BA/ B.COM

Spring, 2025

Q. 1 a) Interpret the significance level in hypotheses testing with the help of figures.

A hypothesis is an educated prediction that can be tested. You will discover the purpose of a hypothesis then learn how one is developed and written. Examples are provided to aid your understanding, and there is a quiz to test your knowledge.

What Is a Hypothesis?

Imagine you have a test at school tomorrow. You stay out late and see a movie with friends. You know that when you study the night before, you get good grades. What do you think will happen on tomorrow's test?

When you answered this question, you formed a hypothesis. A **hypothesis** is a specific, testable prediction. It describes in concrete terms what you expect will happen in a certain circumstance. Your hypothesis may have been, 'If not studying lowers test performance and I do not study, then I will get a low grade on the test.'

The Purpose of a Hypothesis

A hypothesis is used in an experiment to define the relationship between two **variables**. The purpose of a hypothesis is to find the answer to a question. A formalized hypothesis will force us to think about what results we should look for in an experiment.

The first variable is called the **independent variable**. This is the part of the experiment that can be changed and tested. The independent variable happens first and can be considered the cause of any changes in the outcome. The outcome is called the **dependent variable**. The independent variable in our previous example is not studying for a test. The dependent variable that you are using to measure outcome is your test score.

Let's use the previous example again to illustrate these ideas. The hypothesis is testable because you will receive a score on your test performance. It is measurable because you can compare test scores received from when you did study and test scores received from when you did not study.

A hypothesis should always:

- Explain what you expect to happen
- Be clear and understandable
- Be testable
- Be measurable
- And contain an independent and dependent variable

How to Develop a Hypothesis

Another important aspect of a hypothesis is that it should be based on research. Remember that the purpose of a hypothesis is to find the answer to a question. The first thing you should do if you want to answer a question is to find as much information on the topic as you can. Before you come up with a specific hypothesis, spend some time doing research. Then, start thinking of questions you still have. After thoroughly researching your question, you should have an educated guess about how things work. This guess about the answer to your question is where your hypothesis comes from.

Let's imagine that you want to know why the leaves on the tree in your front yard change color in the fall. First, you would research this phenomenon. You observe what you see happen and read about the subject. You discover that the color change happens when the temperature cools. What question does this information make you ask?

You come up with the following question: 'Does temperature cause the leaves to change color on the tree in my front yard?' Next, you ask yourself if this can be tested. If it can be tested, you'll write a hypothesis that states what you expect to find. Your hypothesis could be 'If lower temperatures cause leaves to change color and the temperature surrounding a tree is decreased, then the leaves will change color.'

How to Write a Hypothesis

Let's learn how to properly write a hypothesis using the previous example of tomorrow's test. Examine the differences in the following hypotheses:

Not studying may cause a lower grade on my test.

This statement is not clear enough to be useful. Your hypothesis should be as specific as possible. You're trying to find the answer to a question. If the hypothesis is vague, it's unclear how to find the answer to your question.

After figuring out what you want to study, what is the next step in designing a research experiment? You, the researcher, write a hypothesis and null hypothesis. This lesson explores the process and terminology used in writing a hypothesis and null hypothesis.

Research Question

After determining a specific area of study, writing a hypothesis and a null hypothesis is the second step in the experimental design process. But before you start writing a hypothesis and a null hypothesis, which we will get to, you have to have a question. This is the bottom, or base, which you will build up from.

What are you interested in? What are you curious about? This is a good place to start because your research should answer the question. Curious about the effects of bright lights on studying. You take the thing you are interested in and turn it into a question. Here is mine: 'What is the effect of bright light on studying?' That's how easy it is to write a research question. Next we will explore how to formulate a research hypothesis based on your research question, then we'll look into what a null hypothesis is and how to write one of these.

Formulating a Hypothesis

You have a question and now you need to turn it into a hypothesis. A **hypothesis** is an educated prediction that provides an explanation for an observed event. An **observed event** is a measurable result or condition. If you can't measure it, then you can't form a hypothesis about it because you can't confirm or reject it. In addition, a hypothesis typically takes the form of an if-then statement so you can test it with your research. What does our hypothesis look like?

'If we increase the amount of light during studying, then the participant's performance on test scores will decrease.'

Let's break down our hypothesis. First off, it is an if-then statement: 'If we increase..., then the participant's...' This creates a prediction that we can test by increasing the light on participants as they study and then see if their test scores go down. It also means that the hypothesis can be proven correct or incorrect based on what happens to the test scores. If test scores don't change, then our hypothesis was incorrect and we will reject it.

You probably also noticed that we changed 'studying' to 'test scores' and the vague term about 'bright light' into 'amount of light.' This is an example of **operationalizing**, which is finding a way to measure or quantify a variable. Studying can't really be researched, but test scores can. And they are basically the same thing since studying typically increases test scores. Also, simply saying 'light' is too vague to be useful or researched, so it was turned into 'amount of light.'

Null Hypothesis

After you formulated your research hypothesis, what if there isn't a connection between light and studying? That is kind of what a null hypothesis is; a **null hypothesis** is defined as a prediction that there will be no effect observed during the study. The reason researchers develop a null hypothesis is to ensure that their research can be proven false. So whenever you are conducting an experiment with a hypothesis, you will create a null hypothesis. Research typically includes a hypothesis, and when this is the case you will form a null hypothesis as a counterbalance to ensure there is a way to disprove your prediction.

Constructing a viable scientific hypothesis involves several different factors. In this lesson, you'll explore what separates a good hypothesis from a bad one and how to identify if your hypothesis has been formulated properly.

Scientific Knowledge

The word 'science' comes from the Latin 'to know.' This aptly describes what scientists do all the time - they ask questions to learn and know more about our natural world. Humans are incredibly curious, and we want to know as much as possible about other organisms and processes that occur in the environment around us. To do this, we ask questions - the who, what, when, where, and why of life.

These questions often lead scientists to develop **hypotheses**, which are proposed explanations for scientific observations. They are not 'educated guesses,' as you may have heard before. Hypotheses are developed very carefully, and they specifically aim to describe and/or explain natural phenomena.

A good hypothesis will lead a scientist to predictions that can then be tested through observations or experimentation. The ability to test a hypothesis is key - if you can't test it, you can't determine if it's valid or not! For example, saying that there is another universe outside of ours is an interesting idea, but you can't test it, so it's not a viable hypothesis.

There are a number of other factors that make hypotheses viable. Not only must hypotheses be testable, they need to be clear statements, should identify measurable variables, and be limited to the experiment at hand.

Hypotheses Follow Questions

You can't have a good hypothesis without a good question. That's like putting the cart before the horse - you're just not going to get anywhere! Your scientific question should identify what you're interested in testing. For example, will feeding plants a certain amount of fertilizer make them grow bigger than plants that do not get fertilizer? Or, do low-oxygen waters increase fish mortality?

Do you see how each of these questions is trying to get at a very specific answer? From such questions, you can devise clear research plans to try and solve your problem. This process starts with a solid hypothesis!

Let's say, for example, that you're trying to answer that first question about plant fertilizer. Your question is specific, and you can easily devise a research experiment to find your answer, so let's formulate a good hypothesis. If you believe that because plants need food to survive that an excess of food might cause increased growth in your plants, your hypothesis could be something like, 'plants that are given fertilizer will grow bigger than plants that are not given fertilizer.'

Does this hypothesis fit our criteria? Let's see, first of all, it is testable. You could run an experiment and feed some plants fertilizer while not giving fertilizer to others. It's also a clear statement - it clearly describes what will happen to the plants when they are given fertilizer. The variables are also measurable - both the amount of fertilizer and the growth of the plants can easily be measured and compared. Finally, it is concise - you are making a statement about something that could be determined in one experiment. You won't need to test this one six different ways to get your answer.

Refining a Hypothesis

That was an example of a good hypothesis, but some hypotheses just don't make the cut. These would be statements that are vague, that can't be tested, and that might take several attempts to satisfy with an answer.

b) Some financial theoreticians believe that the stock market's daily prices constitute a random walk with positive drift. If this is accurate, then the Dow Jones Industrial Average should show a gain on more than 50% of all trading days. If the average increased on 101 to 175 randomly chosen days, what do you think about the suggested theory at 0.01 level of significance? (10+10)

i. $H_o: P \le 0.50$ Ii. $\propto = 0.01$

iii. Test Statistic: $Z = \frac{\hat{P} - P}{\sigma_{\hat{p}}}$

iv. Calculations:

P= 0.50, q= 1-p= 1-0.50= 0.50, x = 101, n = 175, $\hat{P} = \frac{x}{n} = \frac{101}{175} = 0.58$

$$\sigma_{\hat{p}} = \sqrt{\frac{Pq}{n}} = \sqrt{\frac{(0.50)(0.50)}{175}} = 0.04 \quad and \quad Z = \frac{0.58 - 0.50}{0.04} = \frac{0.08}{0.04} = 2$$

v. Critical Region $Z < -2.33$ or $Z > 2.33$

vi. Conclusion: $Z_{cal} = 2 < Z_{table} = 2.33$ $Accept\ H_o$

Q. 2 Two independent samples were collected. For the first sample of 60 observations, the mean was 86 and the standard deviation 6. The second sample of 75 elements had a mean of 82 and a standard deviation of 9. (10+10)

(a) Compute the estimated standard error of the difference between the two means.

$$n_1 = 60, \overline{x_1} = 86, \sigma_1 = 6, n_2 = 75, \overline{x_2} = 82, \sigma_2 = 9$$

a) $$a_{\overline{x2},\overline{x1}} = \sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2}} = \sqrt{\frac{(6)^2}{60} + \frac{(9)^2}{75}} = \sqrt{\frac{36}{60} + \frac{81}{75}} = \sqrt{0.6 + 1.08} = \sqrt{1.68} = 1.30$$

(b) Using a 0.01 level of significance, test whether the two samples can be considered to have come from populations with the same mean.

(a)

i. $H_0: \mu_1 = \mu_2$ $H_0: \mu_1 \ne \mu_2$ ii. $\mu = 0.01$

iii. Test statistics: $Z = \dfrac{\left(\overline{X}_1 - \overline{X}_2\right) - -\left(\mu_1 - \mu_2\right)}{\sigma_{x_1 - x_2}}$

iv. Calculation $n_1 = 60\ \overline{X}_1 = 75, \sigma_1 = 6, n_2 = 75\ \overline{X}_2 = 82, \sigma_2 = 9, \sigma_{X_1, X_2} = 1.30$

$$Z = \frac{(86-82)-(0)}{1.30} = \frac{4}{1.30} = 3.08$$

v. Critical Region: $Z < -2.58$ or $Z > 2.58$

vi. Conclusion $Z_{cal} = 3.08 > Z_{table} = 2.58$ Reject H_o

Q. 3 a) Write properties of the sampling distribution for difference between sample proportions. (10+10)

Sometimes "measuring" or "testing" something destroys it. The government requires automakers who want to sell cars in the U.S. to demonstrate that their cars can survive certain crash tests. Obviously, the company can't be expected to crash every car, to see if it survives! So the company crashes only a sample of cars. Another reason for sampling is that not all units in the population can be identified, such as all the air molecules in the LA basin. So to measure air pollution, you take a sample of air molecules. Also, even if all those air molecules could be identified, it would be too expensive and too time consuming to measure them all.

When conducting research, it is almost always impossible to study the entire population that you are interested in. For example, if you were studying political views among college students in the United States, it would be nearly impossible to survey every single college student across the country. If you were to survey the entire population, it would be extremely timely and costly. As a result, researchers use samples as a way to gather data. A sample is a subset of the population being studied. It represents the larger population and is used to draw inferences about that population. It is a research technique widely used in the social sciences as a way to gather information about a population without having to measure the entire population.

There are several different types and ways of choosing a sample from a population, from simple to complex.

Types of Samples:

Non-probability (non-random) samples:

These samples focus on volunteers, easily available units, or those that just happen to be present when the research is done. Non-probability samples are useful for quick and cheap studies, for case studies, for qualitative research, for pilot studies, and for developing hypotheses for future research.

Reliance on Available Subjects

Relying on available subjects, such as stopping people on a street corner as they pass by, is one method of sampling, although it is extremely risky and comes with many cautions. This method, sometimes referred to as a convenience sample, does not allow the researcher to have any control over the representativeness of the sample. It is only justified if the researcher wants to study the characteristics of people passing by the street corner at a certain point in time or if other sampling methods are not possible. The researcher must also take caution to not use results from a convenience sample to generalize to a wider population.

Purposive or Judgmental Sample

A purposive, or judgmental, sample is one that is selected based on the knowledge of a population and the purpose of the study. For example, if a researcher is studying the nature of school spirit as exhibited at a school pep rally, he or she might interview people who did not appear to be caught up in the emotions of the crowd or students who did not attend the rally at all. In this case, the researcher is using a purposive sample because those being interviewed fit a specific purpose or description.

Snowball Sample.

A snowball sample is appropriate to use in research when the members of a population are difficult to locate, such as homeless individuals, migrant workers, or undocumented immigrants. A snowball sample is one in which the researcher collects data on the few members of the target population he or she can locate, then asks those individuals to provide information needed to locate other members of that population whom they know. For example, if a researcher wishes to interview undocumented immigrants from Mexico, he or she might interview a few undocumented individuals that he or she knows or can locate and would then rely on those subjects to help locate more undocumented individuals. This process continues until the researcher has all the interviews he or she needs or until all contacts have been exhausted.

Quota Sample

A quota sample is one in which units are selected into a sample on the basis of pre-specified characteristics so that the total sample has the same distribution of characteristics assumed to exist in the population being studied. For example, if you a researcher conducting a national quota sample, you might need to know what proportion of the population is male and what proportion is female as well as what proportions of each gender fall into different age categories, race or ethnic categories, educational categories, etc. The researcher would then collect a sample with the same proportions as the national population.

Probability Sampling Techniques

Probability sampling is a sampling technique where the samples are gathered in a process that gives all the individuals in the population equal chances of being selected.

Simple Random Sample

The simple random sample is the basic sampling method assumed in statistical methods and computations. To collect a simple random sample, each unit of the target population is assigned a number. A set of random numbers is then generated and the units having those numbers are included in the sample. For example, let's say you have a population of 1,000 people and you wish to choose a simple random sample of 50 people. First, each person is numbered 1 through 1,000. Then, you generate a list of 50 random numbers (typically with a computer program) and those individuals assigned those numbers are the ones you include in the sample.

Systematic Sample

In a systematic sample, the elements of the population are put into a list and then every kth element in the list is chosen (systematically) for inclusion in the sample. For example, if the population of study contained 2,000 students at a high school and the researcher wanted a sample of 100 students, the students would be put into list form and then every 20th student would be selected for inclusion in the sample. To ensure against any possible human bias in this method, the researcher should select the first individual at random. This is technically called a systematic sample with a random start.

Stratified Sample

A stratified sample is a sampling technique in which the researcher divided the entire target population into different subgroups, or strata, and then randomly selects the final subjects proportionally from the different strata. This type of sampling is used when the researcher wants to highlight specific subgroups within the population. For example, to obtain a stratified sample of university students, the researcher would first organize the population by college class and then select appropriate numbers of freshmen, sophomores, juniors, and seniors. This ensures that the researcher has adequate amounts of subjects from each class in the final sample.

Cluster Sample

Cluster sampling may be used when it is either impossible or impractical to compile an exhaustive list of the elements that make up the target population. Usually, however, the population elements are already grouped into subpopulations and lists of those subpopulations already exist or can be created. For example, let's say the target population in a study was church members in the United States. There is no list of all church members in the country. The researcher could, however, create a list of churches in the United States, choose a sample of churches, and then obtain lists of members from those churches.

Systematic random sampling:

Each unit in the population is identified, and each unit has an equal chance of being in the sample.

For example, to select a sample of 25 dorm rooms in your college dorm, make a list of all the room numbers in the dorm. Say there are 100 rooms. Divide the total number of rooms (100) by the number of rooms you want in the sample (25). The answer is 4. This means that you are going to select every fourth dorm room from the list. But you must first consult a table of random numbers. Pick any point on the table, and read across or down until you come to a number between 1 and 4. This is your random starting point. Say your random starting point is "3". This means you select dorm room 3 as your first room, and then every fourth room down the list (3, 7, 11, 15, 19, etc.) until you have 25 rooms selected.

This method is useful for selecting large samples, say 100 or more. It is less cumbersome than a simple random sample using either a table of random numbers or a lottery method. For example, you might have to sample files in a large filing cabinet. It is easier to select every 17th file than to pull out all the files and number them, etc.

However, you must be aware of problems that can arise in systematic random sampling. If the selection interval matches some pattern in the list (e.g., each 4th dorm room is a single unit, where all the others are doubles) you will introduce systematic bias into your sample.

Stratified random sampling:

Each unit in the population is identified, and each unit has a known, non-zero chance of being in the sample. This is used when the researcher knows that the population has sub-groups (strata) that are of interest.

For example, if you wanted to find out the attitudes of students on your campus about immigration, you may want to be sure to sample students who are from every region of the country as well as foreign students. Say your student body of 10,000 students is made up of 8,000 - West; 1,000 - East; 500 - Midwest; 300 - South; 200 - Foreign.

If you select a simple random sample of 500 students, you might not get any from the Midwest, South, or Foreign. To make sure that you get some students from each group, you can divide the students into these five groups, and then select the same percentage of students from each group using a simple random sampling method. This is proportional stratified random sampling.

However, you may still have too few of some types of students. Instead, you may divide students into the five groups and then select the same number of students from each group using a simple random sampling method. This is disproportionate stratified random sampling. This allows you to have enough students in each sub-group so that you can perform some meaningful statistical analyses of the attitudes of students in each sub-group. In order to say something about the attitudes of the total student population of the university, however, you will have to apply weights to the findings for each sub-group, proportional to its presence in the total student body.

b) A sample of 32 money-market mutual funds was chosen on January 1, 1996, and the average annual rate of return over the past 30 days was found to be 3.23% and the sample standard deviation was 0.51%. A year earlier, a sample of 38 money-market funds showed an average rate of return of 4.36% and the sample standard deviation was 0.84%. Is it reasonable to conclude (at α = 0.05) that money-market interest rates declined during 1995?

i. $H_0: \mu_1 \ge \mu_2$ or $H_0: \mu_1 - \mu_2 \ge 0$

$H_1: \mu_1 < \mu_2$ $H_1: \mu_1 - \mu_2 < 0$

ii. $\sigma = 0.05$

iii. Test Statistics $Z = \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}}$

iv. Calculation = $n_2 = 38$, $\bar{x}_2 = 4.36$, $\hat{\sigma}_2 = 0.84$

$n_2 = 32$, $\bar{x}_1 = 3.23$, $\hat{\sigma}_1 = 0.51$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} = \sqrt{\frac{(0.51)^2}{32} + \frac{(0.84)^2}{38}} = 0.16$$

$$Z = \frac{(3.23 - 4.36) - 0}{0.16} = \frac{1.13}{0.16} = -7.06$$

v. Critical Region: $Z < -1.645$

vi. Conclusion: $Z_{cal} = -7.06 > Z_{table} = -1.645$

Reject Ho

Q. 4 a) What is a linear regression model? Explain the assumptions underlying the linear regression model. (10+10)

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a "least squares" method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).

Generate predictions more easily

You can perform linear regression in Microsoft Excel or use statistical software packages such as IBM SPSS® Statistics that greatly simplify the process of using linear-regression equations, linear-regression models and linear-regression formula. SPSS Statistics can be leveraged in techniques such as simple linear regression and multiple linear regression.

You can perform the linear regression method in a variety of programs and environments, including:

- R linear regression.
- MATLAB linear regression.
- Sklearn linear regression.
- Linear regression Python.
- Excel linear regression.

Why linear regression is important

Linear-regression models are relatively simple and provide an easy-to-interpret mathematical formula that can generate predictions. Linear regression can be applied to various areas in business and academic study.

You'll find that linear regression is used in everything from biological, behavioral, environmental and social sciences to business. Linear-regression models have become a proven way to scientifically and reliably predict the future. Because linear regression is a long-established statistical procedure, the properties of linear-regression models are well understood and can be trained very quickly.

IBM TechXchange Conference 2024 | October 21-24 in Las Vegas

Join the must-attend event for technologists using IBM products and solutions. Explore the growing session catalog of over 1,200 sessions and labs.
Explore and register for TechXchange

What is linear regression?

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a "least squares" method to discover the best-fit line for a set of paired data. You then estimate the value of X (dependent variable) from Y (independent variable).

Discover the power of integrating a data lakehouse strategy into your data architecture, including enhancements to scale AI and cost optimization opportunities.
Related content

Register for the IDC report
Generate predictions more easily

You can perform linear regression in Microsoft Excel or use statistical software packages such as IBM SPSS® Statistics that greatly simplify the process of using linear-regression equations, linear-regression models and linear-regression formula. SPSS Statistics can be leveraged in techniques such as simple linear regression and multiple linear regression.

You can perform the linear regression method in a variety of programs and environments, including:

- R linear regression.
- MATLAB linear regression.
- Sklearn linear regression.
- Linear regression Python.
- Excel linear regression.

Why linear regression is important

Linear-regression models are relatively simple and provide an easy-to-interpret mathematical formula that can generate predictions. Linear regression can be applied to various areas in business and academic study.

You'll find that linear regression is used in everything from biological, behavioral, environmental and social sciences to business. Linear-regression models have become a proven way to scientifically and reliably predict the future. Because linear regression is a long-established statistical procedure, the properties of linear-regression models are well understood and can be trained very quickly.
A proven way to scientifically and reliably predict the future

Business and organizational leaders can make better decisions by using linear regression techniques. Organizations collect masses of data, and linear regression helps them use that data to better manage reality — instead of relying on experience and intuition. You can take large amounts of raw data and transform it into actionable information.

You can also use linear regression to provide better insights by uncovering patterns and relationships that your business colleagues might have previously seen and thought they already understood. For example, performing an analysis of sales and purchase data can help you uncover specific purchasing patterns on particular days or at certain times. Insights gathered from regression analysis can help business leaders anticipate times when their company's products will be in high demand.

Assumptions to be considered for success with linear-regression analysis:

- **For each variable**: Consider the number of valid cases, mean and standard deviation.
- **For each model**: Consider regression coefficients, correlation matrix, part and partial correlations, multiple R, R2, adjusted R2, change in R2, standard error of the estimate, analysis-of-variance table, predicted values and residuals. Also, consider 95-percent-confidence intervals for each regression coefficient, variance-covariance matrix, variance inflation factor, tolerance, Durbin-Watson test, distance measures (Mahalanobis, Cook and leverage values), DfBeta, DfFit, prediction intervals and case-wise diagnostic information.
- **Plots**: Consider scatterplots, partial plots, histograms and normal probability plots.

- **Data**: Dependent and independent variables should be quantitative. Categorical variables, such as religion, major field of study or region of residence, need to be recoded to binary (dummy) variables or other types of contrast variables.
- **Other assumptions**: For each value of the independent variable, the distribution of the dependent variable must be normal. The variance of the distribution of the dependent variable should be constant for all values of the independent variable. The relationship between the dependent variable and each independent variable should be linear and all observations should be independent.

b) The following data shows the son's height and father's height.

Father height (X)	**59**	**61**	**63**	**65**	**67**	**69**	**71**	**73**	**75**
Son's height (Y)	**64**	**66**	**67**	**67**	**68**	**69**	**70**	**72**	**72**

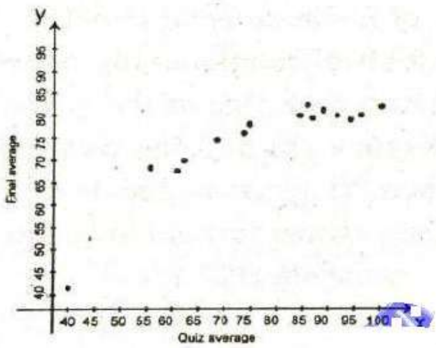
Estimate the regression line y on x.

Q4.

Solution:

(a) Dependent (y) Variable = Final Average
Independent (x) Variable = Quiz Average

(b)



(c) Linear relationship

(d) Yes, the final average of the students depends upon their quiz average.

Q. 5 a) Discuss different types of index numbers in some detail. (10+10)

Meaning of Index Numbers 2. Features of Index Numbers 3. Steps or Problems in the Construction 4. Construction of Price Index Numbers (Formula and Examples) 5. Difficulties in Measuring Changes in Value of Money 6. Types of Index Numbers 7. Importance 8. Limitations.

Meaning of Index Numbers:

The value of money does not remain constant over time. It rises or falls and is inversely related to the changes in the price level. A rise in the price level means a fall in the value of money and a fall in the price level means a rise in the value of money. Thus, changes in the value of money are reflected by the changes in the general level of prices over a period of time. Changes in the general level of prices can be measured by a statistical device known as 'index number.

Index number is a technique of measuring changes in a variable or group of variables with respect to time, geographical location or other characteristics. There can be various types of index numbers, but, in the present context, we are concerned with price index numbers, which measures changes in the general price level (or in the value of money) over a period of time.

Price index number indicates the average of changes in the prices of representative commodities at one time in comparison with that at some other time taken as the base period. According to L.V. Lester, "An index number of prices is a figure showing the height of average prices at one time relative to their height at some other time which is taken as the base period.

Features of Index Numbers:

The following are the main features of index numbers:

(i) Index numbers are a special type of average. Whereas mean, median and mode measure the absolute changes and are used to compare only those series which are expressed in the same units, the technique of index numbers is used to measure the relative changes in the level of a phenomenon where the measurement of absolute change is not possible and the series are expressed in different types of items.
(ii) Index numbers are meant to study the changes in the effects of such factors which cannot be measured directly. For example, the general price level is an imaginary concept and is not capable of direct measurement. But, through the technique of index numbers, it is possible to have an idea of relative changes in the general level of prices by measuring relative changes in the price level of different commodities.

(iii) The technique of index numbers measures changes in one variable or group of related variables. For example, one variable can be the price of wheat, and group of variables can be the price of sugar, the price of milk and the price of rice.

(iv) The technique of index numbers is used to compare the levels of a phenomenon on a certain date with its level on some previous date (e.g., the price level in 1980 as compared to that in 1960 taken as the base year) or the levels of a phenomenon at different places on the same date (e.g., the price level in India in 1980 in comparison with that in other countries in 1980).

Steps or Problems in the Construction of Price Index Numbers:

The construction of the price index numbers involves the following steps or problems:
1. Selection of Base Year:

The first step or the problem in preparing the index numbers is the selection of the base year. The base year is defined as that year with reference to which the price changes in other years are compared and expressed as percentages. The base year should be a normal year.

In other words, it should be free from abnormal conditions like wars, famines, floods, political instability, etc. Base year can be selected in two ways- (a) through fixed base method in which the base year remains fixed; and (b) through chain base method in which the base year goes on changing, e.g., for 1980 the base year will be 1979, for 1979 it will be 1978, and so on.

2. Selection of Commodities:

The second problem in the construction of index numbers is the selection of the commodities. Since all commodities cannot be included, only representative commodities should be selected keeping in view the purpose and type of the index number.

In selecting items, the following points are to be kept in mind:

(a) The items should be representative of the tastes, habits and customs of the people.
(b) Items should be recognizable,
(c) Items should be stable in quality over two different periods and places.
(d) The economic and social importance of various items should be considered

Collection of Prices:

After selecting the commodities, the next problem is regarding the collection of their prices:
(a) From where the prices to be collected;
(b) Whether to choose wholesale prices or retail prices;
(c) Whether to include taxes in the prices or not etc.

While collecting prices, the following points are to be noted:
(a) Prices are to be collected from those places where a particular commodity is traded in large quantities.

(b) Published information regarding the prices should also be utilised,
(c) In selecting individuals and institutions who would supply price quotations, care should be taken that they are not biased.
(d) Selection of wholesale or retail prices depends upon the type of index number to be prepared. Wholesale prices are used in the construction of general price index and retail prices are used in the construction of cost-of-living index number.
ADVERTISEMENTS:

(e) Prices collected from various places should be averaged.

4. Selection of Average:

Since the index numbers are, a specialised average, the fourth problem is to choose a suitable average. Theoretically, geometric mean is the best for this purpose. But, in practice, arithmetic mean is used because it is easier to follow.

5. Selection of Weights:

Generally, all the commodities included in the construction' of index numbers are not of equal importance. Therefore, if the index numbers are to be representative, proper weights should be assigned to the commodities according to their relative importance.

For example, the prices of books will be given more weightage while preparing the cost-of-living index for teachers than while preparing the cost-of-living index for the workers. Weights should be unbiased and be rationally and not arbitrarily selected.

6. Purpose of Index Numbers:

The most important consideration in the construction of the index numbers is the objective of the index numbers. All other problems or steps are to be viewed in the light of the purpose for which a particular index number is to be prepared. Since, different index numbers are prepared with specific purposes and no single index number is 'all purpose' index number, it is important to be clear about the purpose of the index number before its construction.

7. Selection of Method:

The selection of a suitable method for the construction of index numbers is the final step.

There are two methods of computing the index numbers:

(a) Simple index number and
(b) Weighted index number.
Simple index number again can be constructed either by – (i) Simple aggregate method, or by (ii) simple average of price relative's method. Similarly, weighted index number can be constructed either by (i) weighted aggregative method, or by (ii) weighted average of price relative's method. The choice of method depends upon the availability of data, degree of accuracy required and the purpose of the study.

Construction of Price Index Numbers (Formula and Examples):
Construction of price index numbers through various methods can be understood with the help of the following examples:

1. Simple Aggregative Method:

In this method, the index number is equal to the sum of prices for the year for which index number is to be found divided by the sum of actual prices for the base year.

b) Discuss the major problems involved in the construction of index numbers of prices.

The construction of index numbers of prices involves several challenges and problems that can significantly affect their accuracy and reliability. Here are the major issues:

1. Selection of Base Year

- The choice of a base year significantly influences the index number. A base year must be representative of typical economic conditions; if it is too unusual, the index may provide misleading information. Economic conditions, inflation rates, and other factors can change over time, affecting the comparability.

2. Selection of Items

- Determining which goods and services to include in the index can be problematic. A representative sample is crucial, but the following issues may arise:

  - **Variety of Goods**: There are numerous goods and services, and selecting which ones to include to accurately represent consumer behavior is challenging.
  - **Changing Consumption Patterns**: Consumer preferences change over time, affecting which items should be included or excluded.

3. Weighting of Items

- The weighting of items in the index reflects their relative importance in consumer spending. Defining appropriate weights can be difficult:

- **Dynamic Spending Behavior**: Unlike static weights, consumer spending patterns may change, leading to inaccuracies if weights are not updated regularly.
- **Assigning Weights**: The selection process for assigning weights involves subjective judgments and can introduce bias.

4. Quality Changes

- Improvements or deteriorations in the quality of goods and services can distort price comparisons:
- For example, if a computer is upgraded with new features, its higher price may not reflect inflation but rather improved quality. Adjusting for quality changes is complex and often requires judgment calls that may lack transparency.

5. Temporal Issues

- Differences in price changes over time can affect index numbers. Seasonal fluctuations, temporary shortages, or surpluses can cause irregular price movements:
- **Seasonal Adjustments**: Proper adjustments may not always be made, which can lead to inaccurate readings of price trends.

6. Geographical Variations

- Prices can differ significantly across regions due to local economic conditions, taxes, and demand. Constructing a national index may not accurately reflect local price levels.
- **Regional Representation**: Ensuring adequate representation of diverse geographical areas in the index is a challenge, and oversights can mislead policymakers and researchers.

7. Data Availability and Reliability

- Availability and reliability of data can be significant issues:
- **Data Source Quality**: The accuracy of the index depends heavily on the quality of data from various sources, which can differ in methodologies and reliability.
- **Timeliness of Data**: Outdated data can result in indices that do not accurately represent current price levels.

8. Different Index Formulas

- The choice of index number formula (e.g., Laspeyres, Paasche, Fisher) impacts the result:
- **Formula Bias**: Different formulas can yield different index numbers, leading to issues of consistency and comparability if not well justified.

9. Interpretation Issues

- Index numbers can be misinterpreted, leading to misguided economic policies or business decisions.
- **Context Sensitivity**: Understanding what an index number signifies requires context, and failure to appreciate this can lead to incorrect conclusions.

10. Disparate Price-Inducing Factors

- Economic variables such as supply chain disruptions, changes in demand, or macroeconomic policies can influence prices unpredictably, leading to errant index figures.

Conclusion

The construction of price index numbers, while essential for economic analysis, faces several challenges that must be navigated carefully to ensure reliability and accuracy. Addressing these issues often requires extensive data collection, statistical expertise, and clear methodologies to provide meaningful and actionable insights.